

SOEN 363: Data Systems for Software Engineers

Assignment 2, Fall 2023

October 24, 2023

Date posted: Tuesday, October 24th, 2023.

Date due: Monday, November 6th, 2023, by 23:59.

Weight: 5% of the overall grade.

Individual assignment. You must work strictly on your own.

Overview

In this assignment, you create a local database of movies and their information. The assignment targets writing SQL queries for data definition, data collection, as well as search queries.

While IMDB [1] (or any other movie public database) may be used as the main source of the data, in this assignment you are required to create your own database and populate the data. The data population may be done through public APIs [2] such as IMDbOT [3].

Most APIs return data in JSON format. Using a JSON Visualizer [5] would be extremely helpful in locating the info items that are returned by the API.

GET `https://search.imdbot.workers.dev/?tt=tt0068646`

You may optionally start with some pre-populated database (see ref. [4]), however, your database must strictly follow the following design.

Movies are uniquely identified by their associated imdb-ids. For each movie, store: title, description / plot, rating, content rating, genres, actors, directors, creators, release year, runtime, AKAs, countries, languages, keywords, and number of reviews.

Note that both IMDbOT and the above Sample Movies Database are given as examples. You may find alternatives on the internet. You are allowed to use any data source that fits the purpose.

Implementation Platform

In this assignment, you may use either MySQL [6] or PostgreSQL [7] to implement the database tables. Clearly indicate the RDBMS system that you use.

SQL Queries

Provide answer to each of the following parts in separate `.sql` files.

Part 1. [15 pts] Provide the DDL queries for creating the database tables that correspond entities and their relationships.

- Use internal integer primary keys.
- Include referential integrity / unique constraints.
- Define a person table and use IS-A relation to address actors, directors, and creators. Note that actors include character names.
- Include full country name and short code for countries.
- Create a foreign tables for genres, content rating, and keywords.
- Use singular names for tables / relations.

Part 2. [25 pts] Provide the DML queries that populates the data in the database. Make sure you create at least 50 movies. While you may use any approach to populate the data into your database, you need to export your database tables into insert queries. Note that the graders will run the DDL queries in the previous section to create your empty database

followed by the insert queries in this section. Make sure the queries are provided in correct order.

Part 3. [10 pts] Create the following views:

- movie-summary: that displays imdb key, title, description / plot, rating, content rating, runtime, number of keywords, number of countries.
- people: imdb key, name, is-actor (yes/no), is-director (yes/no), is-creator (yes/no), total-number-of-movies.

Part 4. [50 pts] Provide the following queries:

- A) **[5 pts]** Find total number of movies in the database.
- B) **[5 pts]** Pick an actor. Find all movies by that actor that is release prior to 2010. List imdb id, movie title, release date,
- C) **[5 pts]** Find movies that have highest number of keywords. Provide the query with and without using the view.
- D) **[5 pts]** Find number of movies that are in more than one language.
- E) **[5 pts]** For each language list how many movies are there in the database. Order by highest rank.
- F) **[5 pts]** List all French Canadian movies.
- G) **[5 pts]** List all non-English movies that are not made in Canada / USA.
- H) **[5 pts]** List all movies that are directed by one of its actors.
- I) **[5 pts]** Find top 3 movies (higher ratings). Sort by rating as well as number of reviews.
- J) **[5 pts]** Find top 3 movies that have at least 10 reviews.

Make sure all above queries return data. Modify the data in your database, if necessary.

Submit your assignment electronically on Moodle: <https://moodle.concordia.ca>

Include your name and student ID in the submission. Make sure that you upload the assignment to the correct assignment box on Moodle. No email submissions are accepted. Assignments uploaded to the wrong system, wrong folder, or submitted via email will be discarded and no resubmission will be allowed. Make sure you can access Moodle prior to the submission deadline. The deadline will not be extended.

References

1. <https://www.imdb.com/>
2. <https://github.com/public-apis/public-apis>
3. <https://github.com/SpEcHiDe/IMDbOT>
4. <https://www.databasestar.com/sample-database-movies/>
5. <https://jsonviewer.stack.hu/>
6. <https://www.mysql.com/>
7. <https://www.postgresql.org/>